Image Interpretation/Mapping
Accuracy Assessment
Thematic Map (Classification) Accuracy Assessment

Information derived from remotely sensed data are important for environmental models at local, regional, and global scales. The remote sensing–derived thematic information may be in the form of thematic maps or statistics derived from area-frame sampling techniques.

The thematic information must be accurate because important decisions are made throughout the world using the information.
Thematic Map (Classification) Accuracy Assessment

Unfortunately, the thematic information contains error.

Image analysts who create remote sensing–derived thematic information should recognize the sources of the error, minimize it as much as possible, and inform the user how much confidence they should have in the thematic information.

Remote sensing–derived thematic maps should normally be subjected to a thorough accuracy assessment before being used in scientific investigations and policy decisions.
Thematic (Classification) Accuracy Assessment

Accuracy assessment is a general term for comparing the classification to geographical data that are assumed to be true, in order to determine the accuracy of the classification process.

Reference Pixels
are the points on the classified image for which actual data are (or will be) known.
**Ground Reference Test Pixels**

The ideal situation is to locate *ground reference test pixels* (or polygons if the classification is based on human visual interpretation) in the study area. These sites are *not* used to train the classification algorithm and therefore represent unbiased reference information.

It is possible to collect some ground reference test information prior to the classification, perhaps at the same time as the training data. But the majority of test reference information is often collected after the classification has been performed using a random sample to collect the appropriate number of unbiased observations per category.

**Sample Size**

Difficult to find an actual number of pixels to get a statistically sound and practicably attainable accuracy assessment.

Good rule of thumb is to collect a minimum of 50 samples for each land-cover category. If the area is especially large or classification system contains more categories (over 12), the minimum number of samples should be increased to 75 or 100 per class.
Sampling Strategy:

Random:
No rules to be used for collecting random points.

Stratified random:
The number of random points to be stratified to the distribution of thematic layer classes.

Equalized Stratified random
Each class to have an equal number of random points.
Evaluation of Error Matrices

After the ground reference test information has been collected from the randomly located sites, the test information is compared pixel by pixel (or polygon by polygon when the remote sensor data are visually interpreted) with the information in the remote sensing–derived classification map.

Agreement and disagreement are summarized in the cells of the error matrix. Information in the error matrix may be evaluated using simple descriptive statistics or multivariate analytical statistical techniques.

Error Matrix

Error Matrix is a square array of numbers laid out in rows and columns that expresses the number of sample units assigned to a particular category relative to the actual category as verified in the field. The columns normally represent the reference data, while the rows indicate the classification generated from remotely sensed data.

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>k</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x_{1,1}$</td>
<td>$x_{1,2}$</td>
<td>$x_{1,3}$</td>
<td>$x_{1,k}$</td>
<td>$x_{1,}$</td>
</tr>
<tr>
<td>2</td>
<td>$x_{2,1}$</td>
<td>$x_{2,2}$</td>
<td>$x_{2,3}$</td>
<td>$x_{2,k}$</td>
<td>$x_{2,}$</td>
</tr>
<tr>
<td>3</td>
<td>$x_{3,1}$</td>
<td>$x_{3,2}$</td>
<td>$x_{3,3}$</td>
<td>$x_{3,k}$</td>
<td>$x_{3,}$</td>
</tr>
<tr>
<td>k</td>
<td>$x_{k,1}$</td>
<td>$x_{k,2}$</td>
<td>$x_{k,3}$</td>
<td>$x_{k,k}$</td>
<td>$x_{k,}$</td>
</tr>
<tr>
<td>Column total</td>
<td>$x_{+1}$</td>
<td>$x_{+2}$</td>
<td>$x_{+3}$</td>
<td>$x_{+k}$</td>
<td>$N$</td>
</tr>
</tbody>
</table>
Descriptive Statistics

The overall accuracy of the classification map is determined by dividing the total correct pixels (sum of the major diagonal) by the total number of pixels in the error matrix (N).

<table>
<thead>
<tr>
<th>Remote Sensing Classification</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Ground Reference Test Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>k</td>
<td>Row total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$x_{1,1}$</td>
<td>$x_{1,2}$</td>
<td>$x_{1,3}$</td>
<td>$x_{1,k}$</td>
<td>$x_{1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$x_{2,1}$</td>
<td>$x_{2,2}$</td>
<td>$x_{2,3}$</td>
<td>$x_{2,k}$</td>
<td>$x_{2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$x_{3,1}$</td>
<td>$x_{3,2}$</td>
<td>$x_{3,3}$</td>
<td>$x_{3,k}$</td>
<td>$x_{3}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>$x_{k,1}$</td>
<td>$x_{k,2}$</td>
<td>$x_{k,3}$</td>
<td>$x_{k,k}$</td>
<td>$x_{k}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column total</td>
<td>$x_{+1}$</td>
<td>$x_{+2}$</td>
<td>$x_{+3}$</td>
<td>$x_{+k}$</td>
<td>$N$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Descriptive Statistics

Computing the accuracy of individual categories is more complex because the analyst has the choice of dividing the number of correct pixels in the category by the total number of pixels in the corresponding row or column.

The total number of correct pixels in a category is divided by the total number of pixels of that category as derived from the reference data (i.e., the column total). This statistic indicates the probability of a reference pixel being correctly classified and is a measure of omission error, or error of exclusion.

This statistic is also called the producer’s accuracy because the producer (the analyst) of the classification is interested in how well a certain area can be classified.
**Producer’s Accuracy**: Measure of *Omission Error* (error of exclusion)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Residential</th>
<th>Commercial</th>
<th>Wetland</th>
<th>Forest</th>
<th>Water</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential</td>
<td>79</td>
<td>3</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td>Commercial</td>
<td>3</td>
<td>55</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>58</td>
</tr>
<tr>
<td>Wetland</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>6</td>
<td>99</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>37</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>Column Total</td>
<td>73</td>
<td>80</td>
<td>103</td>
<td>50</td>
<td>121</td>
<td>407</td>
</tr>
</tbody>
</table>

Overall Accuracy = 382/407 = 94.36%

<table>
<thead>
<tr>
<th>Producer’s Accuracy (measure of omission error)</th>
<th>User’s Accuracy (measure of commission error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential = 70/73 = 96% 4% omission error</td>
<td>Residential = 70/95 = 80% 20% commission error</td>
</tr>
<tr>
<td>Commercial = 55/60 = 92% 8% omission error</td>
<td>Commercial = 55/58 = 95% 5% commission error</td>
</tr>
<tr>
<td>Wetland = 99/105 = 96% 4% omission error</td>
<td>Wetland = 99/93 = 100% 0% commission error</td>
</tr>
<tr>
<td>Forest = 37/50 = 74% 26% omission error</td>
<td>Forest = 37/41 = 90% 10% commission error</td>
</tr>
<tr>
<td>Water = 121/121 = 100% 0% omission error</td>
<td>Water = 121/121 = 100% 0% commission error</td>
</tr>
</tbody>
</table>

**Descriptive Statistics**

If the total number of correct pixels in a category is divided by the total number of pixels that were actually classified in that category, the result is a measure of *commission error*, or error of inclusion.

This measure, called the *user’s accuracy* or reliability, is the probability that a pixel classified on the map actually represents that category on the ground.
User’s Accuracy:
Measure of *Commission Error* (errors of inclusion)

**Kappa Coefficient**

The *Kappa coefficient* expresses the proportionate reduction in error generated by a classification process, compared with the error of a completely random classification.

For example: a value of 0.82 would imply that the classification process was avoiding 82% of the errors that a completely random classification would generate
Field Reference Data Collection

<table>
<thead>
<tr>
<th></th>
<th>Residential</th>
<th>Commercial</th>
<th>Wetland</th>
<th>Forest</th>
<th>Water</th>
<th>Row total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential</td>
<td>70</td>
<td>5</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td>Commercial</td>
<td>5</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>58</td>
</tr>
<tr>
<td>Wetland</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>Forest</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>97</td>
<td>0</td>
<td>101</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td><strong>Column total</strong></td>
<td><strong>73</strong></td>
<td><strong>60</strong></td>
<td><strong>103</strong></td>
<td><strong>50</strong></td>
<td><strong>132</strong></td>
<td><strong>407</strong></td>
</tr>
</tbody>
</table>

**Overall Accuracy = 102/407 = 93.86%**

**Producer’s Accuracy (commission error)**

- Residential = 70/86 = 81.4% 2% omission error
- Commercial = 55/60 = 91.7% 8% omission error
- Wetland = 99/103 = 96.1% 4% omission error
- Forest = 50/57 = 87.7% 12.3% omission error
- Water = 132/132 = 100% 0% omission error

**User’s Accuracy (commission error)**

- Residential = 70/86 = 81.4% 2% commission error
- Commercial = 55/60 = 91.7% 8% commission error
- Wetland = 99/103 = 96.1% 4% commission error
- Forest = 50/57 = 87.7% 12.3% commission error
- Water = 132/132 = 100% 0% commission error

**Computation of Kappa Coefficient of Agreement**

\[
K = \frac{\sum X_{ij} (\frac{1}{n_i} - \frac{1}{n}) - \sum X_i (\frac{1}{n_j} - \frac{1}{n})}{\sum (\frac{1}{n_i} - \frac{1}{n}) - \sum (\frac{1}{n_j} - \frac{1}{n})}
\]

where \( n = 407 \)

\[
\sum X_{ij} = (70 + 55 + 99 + 50 + 132) = 407
\]

\[
\sum X_i = (70 + 55 + 99) = 224
\]

\[
\sum X_j = (50 + 132) = 182
\]

\[
K = \frac{\sum X_{ij} (\frac{1}{n_i} - \frac{1}{n}) - \sum X_i (\frac{1}{n_j} - \frac{1}{n})}{\sum (\frac{1}{n_i} - \frac{1}{n}) - \sum (\frac{1}{n_j} - \frac{1}{n})}
\]

\[
K = \frac{407 - 224 - 182}{224 - 182} = \frac{61}{42} = 1.45
\]

**Overall Kappa = 0.61**
GPS photos
Vegetation
Hydrology
Soils and sediments
Other observable indicators

> 50% Spartina

10 – 50% Spartina

< 10% Spartina
Examples of GPS Georeferenced Field Photos for VFRDB

Cultivated Agriculture

Open Bushland
An error analysis is only possible if the “truth/reference” set is at least one step closer to reality than the remotely sensed product on which the map is based (Congalton, 1991).

In the case of using Landsat data, we considered finer spatial resolution remote sensing data, for example air photos and, and rigorously collected field observations to be “truth/reference” set.

The accuracy of a remotely sensed data product is equally important as the information presented in the product.

Without known accuracy, the product cannot be used reliably, and therefore, has limited applicability.